

MARC-8

The **MARC-8** charset is a MARC standard used in MARC-21 library records.^[1] The MARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form, and they are frequently used in library database systems. The character encoding now known as MARC-8 was introduced in 1968 as part of the MARC format. Originally based on the Latin alphabet, from 1979 to 1983 the JACKPHY initiative expanded the repertoire to include Japanese, Arabic, Chinese, and Hebrew characters (among others), with the later addition of Cyrillic and Greek scripts. If a character is not representable in MARC-8 of a MARC-21 record, then UTF-8 must be used instead. UTF-8 has support for many more characters than MARC-8, which is rarely used outside library data.



Technical details

MARC-8 uses a variant of the ISO-2022 encoding. It uses escape characters to represent characters beyond the 7-bit ASCII range of characters.

It generally uses the same logical BiDi ordering as Unicode.

The combining characters and base characters are in a different order than used in Unicode. The following are some examples. The combining characters are not always stored in reverse order as Unicode normalization. The MARC-21 standard describes the MARC-8 Unicode conversion issues in more detail.

Displayed Character	Unicode <u>NFD</u>	MARC-8
á	a ´	´ a
â	a . ^	^ . a

Code structure

The ISO/IEC 2022 coding specifies a two-layer mapping between character codes and displayed characters. In MARC-8, character codes from the 7-bit ASCII graphic range (0x20–0x7F) are referred to as "Go" codes, while codes from the "high ASCII" range (0xA0–0xFF) are referred to as the "G1" codes. Graphic character sets are designated and invoked by means of a multiple byte escape sequence consisting of the escape character, an Intermediate character sequence, and a Final character in the form ESC *I F*.

The following table shows the intermediate byte after the ESC byte (hexadecimal 1B), and the corresponding ASCII characters.

Intermediate Bytes^[2]

	G0 set				G1 set			
	<u>SBCS</u>		<u>MBCS</u>		<u>SBCS</u>		<u>MBCS</u>	
Normal ISO-2022	28	(24	\$	29)	24 29	\$)
Alternate ISO-2022 (additional 63+16 sets)	2C	,	24 2C	,\$	2D	-	24 2D	\$-

The following table shows the final bytes in hexadecimal and the corresponding ASCII characters after the intermediate bytes.

Final Bytes^[3]

Bytes	Characters	Name	Type	Comment
31	1	Chinese, Japanese, Korean (<u>EACC</u>)	<u>MBCS</u>	
32	2	Basic Hebrew	<u>SBCS</u>	
33	3	Basic Arabic	<u>SBCS</u>	
34	4	Extended Arabic	<u>SBCS</u>	
42	B	Basic Latin (<u>ASCII</u>)	<u>SBCS</u>	
21 45	IE	Extended Latin (<u>ANSEL</u>)	<u>SBCS</u>	The 21(hex) technically is a second byte of the Intermediate segment of this escape sequence.
4E	N	Basic Cyrillic	<u>SBCS</u>	
51	Q	Extended Cyrillic	<u>SBCS</u>	
53	S	Basic Greek	<u>SBCS</u>	

The EACC is the only multibyte encoding of MARC-8, it encodes each CJK character in three ASCII bytes.

For example, to encode the U+4EBA CJK character (人) you will need the following bytes

```
\x1B\x24\x31\x21\x30\x64
```

The \x1B\x24\x31 switches to EACC/CJK, and the \x21\x30\x64 corresponds to the U+4EBA.

Custom set extension

In addition to the ISO-2022 character sets, the following custom sets are available too. The byte designation follows the escape byte (hexadecimal 1B). There is no intermediate byte.

Final Bytes^[4]

Bytes	Characters	Name	Type	Comment
62	b	Subscript set	<u>SBCS</u>	
67	g	Greek Symbol set	<u>SBCS</u>	The alpha, beta, gamma characters normally do not round trip map to Unicode.
70	p	Superscript set	<u>SBCS</u>	
73	s	Basic Latin (<u>ASCII</u>)	<u>SBCS</u>	

References

1. "Character Sets: Introduction: MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media (Library of Congress)" (<https://www.loc.gov/marc/specifications/speccharintro.html>). *Library of Congress*.
2. "Character Sets: MARC-8 Encoding Environment: MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media (Library of Congress)" (<https://www.loc.gov/marc/specifications/speccharmarc8.html>). *Library of Congress*.
3. "Character Sets: MARC-8 Encoding Environment: MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media (Library of Congress)" (<https://www.loc.gov/marc/specifications/speccharmarc8.html>). *Library of Congress*.
4. "Character Sets: MARC-8 Encoding Environment: MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media (Library of Congress)" (<https://www.loc.gov/marc/specifications/speccharmarc8.html>). *Library of Congress*.

External links

- [MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media \(https://www.loc.gov/marc/specifications/speccharintro.html\)](https://www.loc.gov/marc/specifications/speccharintro.html) - The official MARC-8 standard as maintained by the [US Library of Congress](#)

Retrieved from "<https://en.wikipedia.org/w/index.php?title=MARC-8&oldid=1059389598>"

This page was last edited on 9 December 2021, at 04:47 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.